

# GHAJAR EXHIBIT 63

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 1

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION

---

Richard Kadrey, et al.,	)	
Individual and Representative	)	
Plaintiffs,	)	CASE NO.
	)	3:23-cv-03417-VC
-against-	)	
	)	
Meta Platforms, Inc.,	)	
Defendant.	)	
	)	

---

\*\*\* HIGHLY CONFIDENTIAL \*\*\*

ATTORNEYS' EYES ONLY

VIDEO-RECORDED DEPOSITION OF  
MELANIE KAMBADUR

Cooley, LLP  
55 Hudson Yards  
New York, New York 10001

09/17/2024  
9:07 a.m. (EDT)

REPORTED BY: MONIQUE CABRERA

---

DIGITAL EVIDENCE GROUP  
1730 M Street, NW, Suite 812  
Washington, D.C. 20036  
(202) 232-0646

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 32

1 MR. WEINSTEIN: Object to form.

2 Vague.

3 A. What do you mean by "important"?

4 BY MR. YOUNG:

5 Q. Have you selected training data that  
6 is fed into a model?

7 A. By "selected" do you mean  
8 specifically -- what -- yeah.

9 What do you mean by "selecting"?

10 Q. So does the training data used to  
11 develop a model affect how a model would perform?

12 A. Yes.

13 Q. So you wouldn't feed just any data  
14 into a model, use my data to train a model,  
15 right?

16 A. Someone could.

17 Q. But you wouldn't, right?

18 A. I guess it dep- -- would depend on  
19 the purpose. But generally, we do select  
20 specific data, yes.

21 Q. High quality data?

22 A. Generally, yes.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 33

1 Q. What do you -- what do you  
2 understand me to mean when I say "high quality  
3 data"?

4 A. Actually, I'm not 100 percent sure  
5 what you mean by high quality data.

6 Q. Is "high quality data" a term you  
7 use in your work as AI developer?

8 A. Yes, we have described data as high  
9 quality levels. But I'm not sure if you have the  
10 same definition as me.

11 Q. Okay. What is your definition of  
12 high quality data?

13 A. Generally, data which at least  
14 small-scale experiments, has been shown to  
15 improve some metric that we're targeting for  
16 downstream model performance.

17 Q. So would high quality data generally  
18 be accurate data?

19 A. It could be.

20 Q. Free of errors, would high quality  
21 data be free of errors?

22 A. Probably. It depends on your goals.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 35

1 training large language models?

2 A. Not directly. I have managed people  
3 who were doing the evaluation and I sometimes  
4 reviewed their evaluations.

5 Q. So by "review evaluations," do you  
6 mean someone who reported to you, would evaluate  
7 a dataset for use in training and then you would  
8 say yes, no, this would be something we would  
9 include?

10 A. I did not solely make the yes/no  
11 decisions. The first part of that statement,  
12 that people who report to me would run  
13 experiments and evaluate models is correct.

14 Q. Who else -- so you said you were not  
15 the person who solely made the decision to  
16 include certain datasets.

17 Who else would have input into that  
18 decision?

19 A. Generally, we looked at the actual  
20 perf- -- the performance of different experiments  
21 and for which -- like, are you asking about a  
22 specific model?

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 36

1 Q. No. Just generally.

2 A. Generally, members of the research  
3 team would look at the experimental outcomes.

4 Q. By "experimental outcomes," these  
5 are quantifiable results, right?

6 A. They are quantitative, but sometimes  
7 noisy metrics that give us an indication of  
8 potential performance.

9 Q. Are these experiments standardized?

10 MR. WEINSTEIN: Object to form.

11 Vague.

12 You can answer.

13 A. I would say that we repeatedly ran  
14 similar experiments but we changed the exact  
15 experimental process from time to time.

16 BY MR. YOUNG:

17 Q. Are you familiar with the Llama  
18 models?

19 A. Yes.

20 Q. What are they?

21 A. They are a series of large language  
22 models that we trained at Meta.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 117

1 dataset that Meta had, but had not processed yet?

2 A. I don't recall when we processed it.

3 Q. Now, do you recall testifying

4 earlier that Libgen was used to pre-train

5 Llama 3, Llama 3 series of models?

6 A. Yes. At least a subset of Libgen

7 was used.

8 Q. But it wasn't used to pre-train the

9 Llama 2 models or the Llama 1 models, right,

10 according to your testimony today?

11 A. Yes, that's correct, to my

12 knowledge.

13 Q. Now, do you remember if Meta had the

14 Libgen dataset before it was processed and

15 including in the Llama 3 models?

16 MR. WEINSTEIN: Object to form.

17 A. Of course we had to have the dataset

18 before including it in the model, yes.

19 BY MR. YOUNG:

20 Q. So that means that there was a

21 decision made to incorporate Libgen, process

22 Libgen as pre-training data for Llama 3, correct?

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 118

1 A. Yes.

2 Q. Do you know who made that decision?

3 A. I don't think there was a  
4 singular -- there was not a singular person.

5 Q. But a decision was made, correct?

6 A. To include Libgen into Llama 3?  
7 Yes.

8 Q. Do you remember why that decision  
9 was made to include Libgen into Llama 3?

10 A. As with all of our datasets, we had  
11 run some small-scale experiments that indicated  
12 it would positively improve our benchmarks on a  
13 larger model. And we had gone through our legal  
14 and privacy review procedures.

15 Q. Did anyone express to you any  
16 concern with using the Libgen data source as  
17 pre-training material?

18 A. I recall people asking if it was  
19 approved.

20 Q. Did you express any concern about  
21 using the Libgen data source as pre-training  
22 material?



9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 337

1 CERTIFICATE OF SHORTHAND REPORTER NOTARY PUBLIC

2 I, Monique Cabrera, the officer  
3 before whom the foregoing deposition was  
4 taken, do hereby certify that the foregoing  
5 transcript is a true and correct record of  
6 the testimony given; that said testimony was  
7 taken by me stenographically and thereafter  
8 reduced to typewriting under my direction;  
9 and that I am neither counsel for, related  
10 to, nor employed by any of the parties to  
11 this case and have no interest, financial or  
12 otherwise, in its outcome.

13 IN WITNESS WHEREOF, I have hereunto  
14 set my hand this 17th day of September, 2024.

15  
16  
17   
18

19 MONIQUE CABRERA

20 Notary Public in and for the State of New York

21 County of Suffolk

My Commission No.

22 Expires: 06/12/2026